

Investigating the Relationship between Sea Surface Temperature and Chlorophyll-*a* Concentration: An Empirical Finding from the North Coast of Semarang, Indonesia

M L Dina^{1*}, M Rohmadoni², V A Firdausa³, A Rusdiyanto⁴, P Y Akshintia⁵, M M Ulkhaq⁶, S Suryanti⁷, A Ghofar⁸, and N. Widyorini⁹

¹ Department of Aquatic Resources, Diponegoro University, Semarang, Indonesia

² Department of Geodetic Engineering, Diponegoro University, Semarang, Indonesia

³ Faculty of Public Health, Diponegoro University, Semarang, Indonesia

⁴ Department of Environmental Engineering, Diponegoro University, Semarang, Indonesia

⁵ PT Pembangunan Jawa-Bali, Surabaya, Indonesia

⁶ Department of Industrial Engineering, Diponegoro University, Semarang, Indonesia

⁷ Department of Aquatic Resources, Diponegoro University, Semarang, Indonesia

⁸ Department of Aquatic Resources, Diponegoro University, Semarang, Indonesia

⁹ Department of Aquatic Resources, Diponegoro University, Semarang, Indonesia

*Corresponding author: mersiliwaudina@students.undip.ac.id

Abstract. This study investigated the relationship between sea surface temperature (SST) and chlorophyll-*a* (chl-*a*) concentration in the North Coast of Semarang, Indonesia. The data were collected using Moderate Resolution Imaging Spectroradiometer Satellite (Aqua-MODIS) chl-*a* level-3 standard mapped image for a period of five years (2015–2019). Due to cloud coverage, monthly averaged data were used in this study. The result shows that monthly averaged SST ranged from 28.1°C to 31.3 °C. This SST is relatively higher in the transition season (especially in the first transition season) compared to the east and west monsoons season. Chl-*a* concentration ranged from 0.002 mg/m³ to 1.388 mg/m³. This concentration fluctuates according to the seasonal winds. The maximum concentration of chl-*a* occurred in the west monsoon season, while the minimum one happened in the second transition season. Using the Pearson correlation coefficient, the correlation is -0.092, which indicates a weak negative correlation between SST and chl-*a* concentration. This study is expected to give an insight into the potential fishing ground since the chl-*a* concentration can be such an indicator for the presence of fish.

1. Introduction

The phytoplankton is a microscopic marine plant that can convert inorganic carbon dioxide into organic carbon through photosynthesis in the upper layers of the ocean. Its concentration (measured as chlorophyll-*a* [chl-*a*] concentration) has been shown to affect top-level predators such as fish [1]. The planktonic ecosystems strip the nutrients such as nitrate, silicate, and phosphate out of the surface layers of the ocean during photosynthesis and hence they also contribute towards the biogeochemical cycling of important chemical elements [2].

Apart from chl-*a* concentration, the existence of small pelagic fishes can also be predicted through physical and biological indicators of the sea surface environment, especially sea surface temperature (SST). It is particularly vulnerable to environmental fluctuations and global change because their short lifespan means that they react rapidly to environmental change [3]; it means that significant change in SST can affect the life of biota in the ocean. Reproduction of fish can also be affected even if SST changes only by 1° to 2° C [4]. Therefore, SST is considered an important factor that regulates the growth of phytoplankton [5].

In order to obtain optimal fishery products, identification of physical and biological conditions in an aquatic ecosystem is necessary; in which SST and chl-*a* regard as important parameters [6]. While chl-*a* concentration provides a measure of enhanced biological production area, SST provides information to explain the oceanic environment suitable for enhanced production [5], [7]. The use of both parameters would improve the understanding of the physical and biological processes of the oceans [8], their productivity [9], and food resource availability for exploring fishery resources [7].

The majority of the ocean's productivity lies within the tropics along with the equatorial band of 10° North Latitude (NL) to 10° South Latitude (SL) [10]. Indonesia, which lies between 6° NL to 11° SL, hence, is categorized as an important fishing area [11] due to its strategic location. One of the potential fishing areas in Indonesia is the North Coast of Java Island, especially the North Coast of Semarang. This study is conducted to identify the relationship between SST and chl-*a* concentration to determine the preferred range of SST and chl-*a* in the North Coast of Semarang. It aims to map the potential fishing ground in that area.

This study utilized Moderate Resolution Imaging Spectroradiometer Satellite (Aqua-MODIS) to capture the data needed. Satellite sensors are able to provide reliable global ocean coverage of SST and chl-*a* at relatively high spatial and temporal resolution. This enables a more efficient analysis of the spatial and temporal distribution that can be measured from space [12]. Direct measurement (e.g., shipboard sampling method) is not an efficient method since it requires expensive cost, time, and limited coverage areas. In addition, MODIS-Aqua image is more accurate because it has a fairly narrow designed wavelength range [13].

2. Data and Methods

This research studied the relationship between SST and chl-*a* concentration in the North Coast of Semarang, Indonesia, which the exact location is 6°49'16.77" SL and 110°22'49.17" East Longitude, see Figure 1. The data were collected using Aqua-MODIS chl-*a* level-3 standard mapped image (SIM) during the period of five years (2015–2019) with monthly intervals and cloud filtering. The data were obtained from National Aeronautics and Space Administration (NASA) Ocean Color (<http://oceancolor.gsfc.nasa.gov>). Chl-*a* concentration values were retrieved from the images using SeaDAS software version 6.1 (provided for ocean colour image processing by NASA). Aqua-MODIS chl-*a* concentration values use the OC2 algorithm for deriving chl-*a* values from the recorded radiance. Sensitivity studies on the algorithm for chl-*a* concentration retrieval from measured sensor detected radiances show that the retrieved chl-*a* values have the radiance error of ~1% [14], [15].

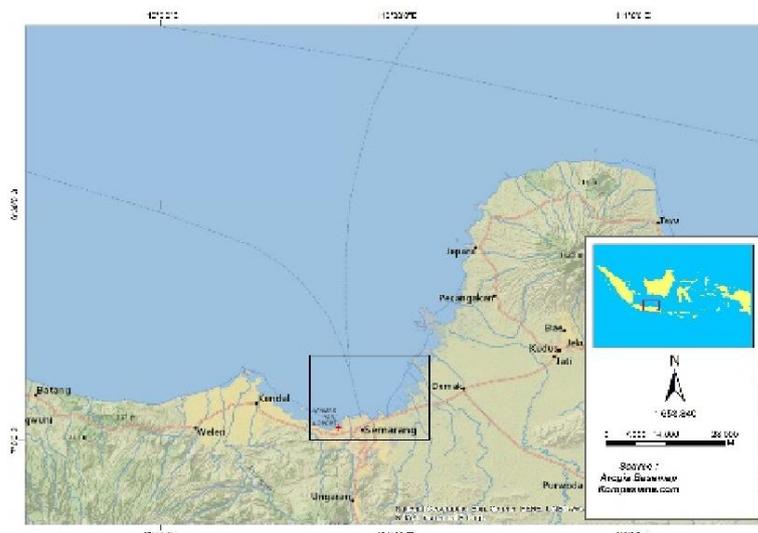


Figure 1. The location of the collected data

Since the collected data is time-series, we modelled the data using the autoregressive integrated moving-average (ARIMA). It is necessary to disentangle the error from the actual data because time-series data often display autocorrelation or serial correlation of the errors across periods [16]. There are three components of the ARIMA model; in which these components can be modified to form any ARIMA model [17], [18], i.e., autoregressive (AR) model, moving average (MA) model, and integrated (I). The general AR model of order p or AR(p) can be written as

$$Z_t = c + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t, \quad (1)$$

where Z_t is a time-series variable, c is a constant, ϕ_p is a parameter of the model (i.e., AR coefficient of order p), ε_t is white noise, and subscript t refers to time notation. The general MA model of order q or MA(q) can be written as

$$Z_t = c + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad (2)$$

where θ_q is a parameter of the model (i.e., MA coefficient of order q). Integrated refers to differencing the data. Note that in the ARIMA procedure, it is necessary to do differencing of the data (i.e., subtracting an observation from observation at the previous time step) to make the time series stationary. For example, when $d = 1$, it means that the data is differenced at the first order, or $Z_t' = Z_t - Z_{t-1}$. In sum, the standard notation of the ARIMA model is ARIMA(p, d, q) where the parameters are substituted with integer values to quickly indicate the specific ARIMA model being used.

There are three stages in modelling the data using ARIMA: (i) identification, (ii) estimation, and (iii) diagnostic checking [19]. In the identification stage, two graphical devices were used to measure the correlation between the observations within a single data series, i.e., autocorrelation function (ACF) and partial autocorrelation function (PACF). The ACF and PACF were used as guides to choose one or more ARIMA models that seem appropriate. The basic idea is that every ARIMA model has a theoretical ACF and PACF associated with it. We then compared the estimated ACF and PACF calculated from the data with various theoretical ACFs and PACFs; and tentatively choose the model whose theoretical ACF and PACF most closely resemble the estimated ACF and PACF of the data. After one (or more) model(s) were tentatively chosen, in the estimation stage, we fit the model(s) to the data to get estimates of ARIMA coefficients (i.e., c , ϕ , and θ). In the diagnostic checking stage, [19] suggested some diagnostic checks to help determine if the model is statistically adequate, e.g., the Ljung-Box test to test whether the residual ACFs are different from zero. Furthermore, the results may also indicate how a model could be improved; this leads us back to the identification stage.

To investigate the relationship between SST and chl- a concentration, the correlation analysis using the Pearson correlation coefficient was used. Notice that we used the data after modelling with ARIMA. The Pearson correlation coefficient is commonly used as a measure of linear correlation between two sets of data. The coefficient is actually the covariance of two variables divided by the product of their standard deviations; thus, it is essentially a normalised measurement of the covariance. Therefore, the coefficient always has a value between -1 and 1 . The sample Pearson correlation coefficient r between two variables (say X and Y) is given by [20]

$$r = \frac{\sum_{i=1}^N (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^N (x_i - \bar{x})^2}} \quad (3)$$

A positive value (or a positive correlation) of r means that when X increases, Y tends to increase; and when X decreases, Y tends to decrease. On the other hand, a negative value (or a negative correlation) of r means that when X increases, Y tends to decrease; and when X decreases, Y tends to increase. A value of 1 implies a perfect positive linear relationship between X and Y ; while a value of -1 implies a perfect negative correlation between X and Y . Finally, a value of 0 implies that there is no linear correlation between X and Y [21]. When the coefficient is closer to 1 (either positive or negative), it indicates a strong relationship between X and Y ; contrarily, when the coefficient is closer to 0, it indicates a weak correlation.

Table 1. Range of sea surface temperature ($^{\circ}$ C)

Season	2015	2016	2017	2018	2019
West Monsoon (Dec. – Feb.)	28.270-28.970	29.512-30.324	28.910-30.734	28.980-29.997	30.628-31.055
First Transition (Mar. – May)	29.100-30.945	30.458-31.114	30.006-30.752	30.113-30.440	30.780-31.379
East Monsoon (June – August)	28.195-29.483	29.939-30.969	28.993-29.852	28.480-29.812	28.603-30.002
Second Transition (Sep. – Nov.)	28.320-30.720	30.331-31.251	29.148-31.292	29.015-30.875	28.729-30.745

Table 2. Range of chlorophyll- a (mg/m^3)

Season	2015	2016	2017	2018	2019
West Monsoon	0.127 – 0.426	0.026 – 0.647	0.620 – 1.388	0.187 – 0.944	0.263 – 0.330
First Transition	0.370 – 0.589	0.341 – 0.467	0.425 – 0.637	0.380 – 0.719	0.605 – 1.065
East Monsoon	0.486 – 0.973	0.450 – 0.693	0.523 – 0.721	0.571 – 0.745	0.378 – 0.708
Second Transition	0.117 – 0.449	0.328 – 0.452	0.002 – 0.448	0.204 – 0.338	0.352 – 0.859

3. Results and Discussion

The range of monthly averaged data during the period of 2015 to 2019 of SST in the North Coast of Semarang were presented in Table 1. In 2015, the lowest SST was in August, i.e., 28.195° C; while the highest SST was in April with 30.945° C. In 2016, the lowest SST was in January, which is 29.512° C; and the highest one was in November, i.e., 31.251° C. Next, in 2017, the lowest SST was in February with 28.910° C; while the highest was in November, which is 31.292° C. In 2018, the lowest one was in August, i.e., 28.480° C and the highest one was in December at 30.977° C. Lastly, in 2019, the lowest and the highest SSTs were in August and April with 28.603° C and 31.379° C, respectively. The low SST is likely due to the upwelling of seawater masses, whereas the high SST is due to the influence of the west monsoon wind. The result indicates that SST decreased in the heating period that occurred from March to August, since in the heating period, solar radiation decreased, resulting in a decrease in short-wave radiation flux [22].

The range of monthly averaged data of chl- a concentration in the North Coast of Semarang during the period of 2015 to 2019 were presented in Table 2. In 2015, the lowest concentration was $0.117 \text{ mg}/\text{m}^3$ where it occurred in October; while the highest one is $0.973 \text{ mg}/\text{m}^3$, occurred in June. In 2016, the lowest figure was $0.026 \text{ mg}/\text{m}^3$ occurred in December; while the highest was $0.694 \text{ mg}/\text{m}^3$ occurred in July. In 2017, the lowest level of chlor- a concentration was $0.002 \text{ mg}/\text{m}^3$ occurred in November and the highest number was $1.388 \text{ mg}/\text{m}^3$ occurred in December. In 2018, the lowest concentration was $0.188 \text{ mg}/\text{m}^3$ happened in January and the highest was in February with the value of $0.945 \text{ mg}/\text{m}^3$. In 2019, the lowest chlor- a concentration was in December with $0.299 \text{ mg}/\text{m}^3$ and the highest was $1.065 \text{ mg}/\text{m}^3$ occurred in March. It seems that the west monsoon and the first transition season are the seasons where the amount of chlor- a concentration is quite high due to high rainfall in Indonesia so that many nutrients come into the sea through the rivers [23].

Before investigating the relationship between SST and chl- a concentration, the previous monthly averaged data were tested for autocorrelation because time-series data is prone

to the presence of autocorrelation. In the identification stage (see Section 2), we used ACF and PACF graphs to see whether there exists autocorrelation in the data. The ACFs and PACFs of SST and chl-*a* concentration are presented in Figure 2. From the ACF and PACF graph, it seems that SST data is autocorrelated since the first lag of ACF as well as the first and the second lags of PACF exceed the upper and lower 95% confidence level. On the other hand, there is no evidence that autocorrelation does present in the chl-*a* concentration data. Next, we have to compare the ACF and PACF graph (of SST data) with various theoretical ACFs and PACFs. In this study, we compared five different ARIMA models, i.e. ARIMA(2,0,0), ARIMA(2,1,1), ARIMA(2,1,0), ARIMA(1,1,1), and ARIMA(1,1,0).

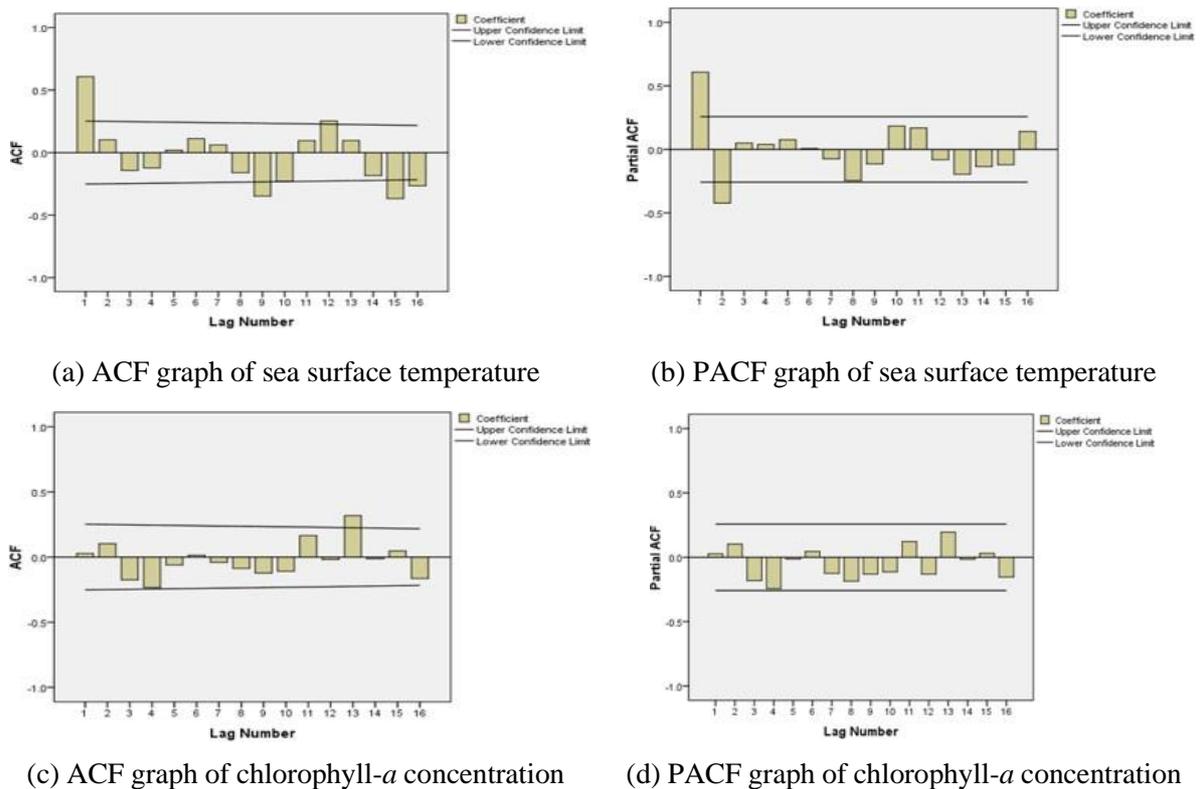


Figure 2. ACF and PACF graphs of sea surface temperature and chlorophyll-*a* concentration

Table 3. Estimation of various ARIMA model's parameters

Parameters	ARIMA(2,0,0)	ARIMA(2,1,1)	ARIMA(2,1,0)	ARIMA(1,1,1)	ARIMA(1,1,0)
Constant	29.948* (0.000)	0.011 (0.247)	0.001 (0.712)	0.003 (0.694)	0.002 (0.691)
ϕ_1	0.933* (0.000)	0.938* (0.000)	0.256* (0.048)	0.000 (0.998)	0.198 (0.132)
ϕ_2	-0.463* (0.000)	-0.445* (0.001)	-0.334* (0.011)	-	-
θ_1	-	0.996 (0.350)	-	1.000 (0.931)	-

*significant at the level of 5%

The estimation of ARIMA parameters was shown in Table 3. According to this estimation, ARIMA(2,0,0) is the most appropriate model since all parameters are statistically significant at the level of 5%. The next stage is diagnostic checking. We tested the residuals of ACF and PACF (see Figure 3) and showed that all ACFs and PACFs do not exceed upper and lower significant level in any lag. After the appropriate ARIMA model is identified, we then forecasted the SST data using ARIMA(2,0,0).

To identify the relationship between SST and chl-*a* concentration, we computed the Pearson correlation coefficient between forecasted SST and chl-*a* concentration data. The result shows that the coefficient is -0.092. This negative correlation is in line with other research, e.g., [15], [24]– [27]. The study by [24] stated that the negative correlation can be explained by the mechanism through wind parameters, meaning that the faster the wind blows, the stronger the mixing process will be. This mixing process will lift the cold air mass with high nutrient content from deeper waters, causing the chl-*a* to concentrate in the surface layer and decreases SST. The study by [25] mentioned when SST reaches its annual maximum in summer, the depth of the mixed layer becomes shallowest. This results in low nutrients supplied to the euphotic layer as vertical mixing is blocked by intensive stratification. As a result, low nutrients limit the growth rate of phytoplankton and result in low chl-*a*. In contrast, a high chl-*a* concentration value in the winter when the mixed layer depth is deeper, and low SST causes weak stratification. Another interesting finding is about the weak correlation result. It is likely because SST and chl-*a* concentration behave differently from each other. This behaviour is due to the presence of wind, upwelling, and other factors such as runoff zone, substrate concentration, temperature gradient, and oxygen abundance.

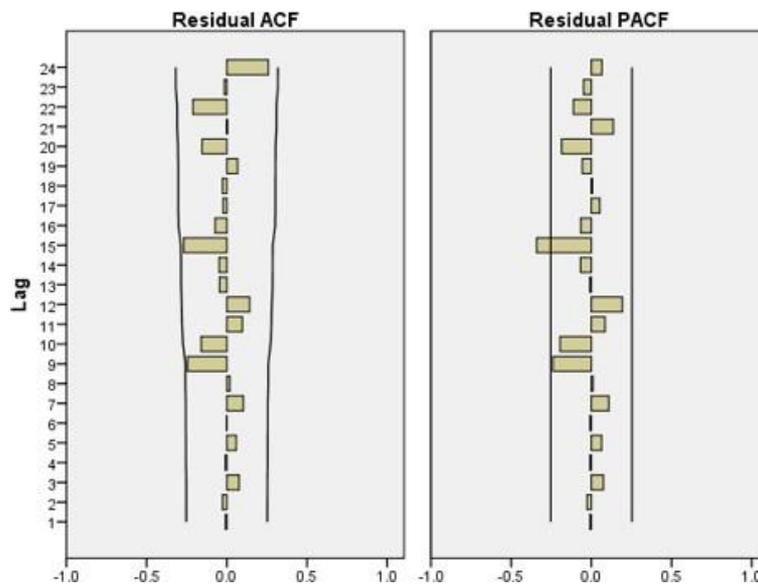


Figure 3. Diagnostic checking with residual ACF and PACF for sea surface temperature data

4. Conclusion

This study investigated the relationship between sea surface temperature (SST) and chlorophyll-*a* (chl-*a*) concentration. Data were collected from 2015-2019 using Aqua-MODIS chl-*a* level-3 SIM with monthly interval and cloud filtering. Data were captured in the North Coast of Semarang, Central Java Province, Indonesia. Because the data is time-series, before investigating the relationship between those two variables, we modelled the data using ARIMA procedure. The result shows that the Pearson correlation coefficient is -0.092. It means the relationship between SST and chl-*a* concentration is negatively and weakly correlated. This result is likely due to several things such as wind and upwelling. For future research, these two factors can be taken into consideration; furthermore, due to the inclusion of several variables, other methods can be considered, such as multiple regression.

5. Acknowledgments

The satellite data used is owned and distributed by the United National Aeronautics and Space Administration. (NASA), supported by the Ocean Biology Processing Group (OBPG) at NASA's Goddard Space Flight Center. The author also expressed invaluable gratitude for input, advice from various parties, especially Ir. Max Rudolf Muskananfolo, M. Sc., Ph. D and Sigit Febrianto, S. Kel., M. Si.

6. References

- [1] Lehodey P, Bertignac M, Hampton J, Lewis A and Picaut J 1997 *Nature* **389** 715–718
- [2] Ravichandran M, Girishkumar M S and Riser S 2012 *Deep-Sea Research I* **65** 15–25
- [3] Cury P and Roy C 1989 *Can. J. Fish. Aquat. Sci.* **46** 670–680
- [4] Wibisana H and Zainab S 2017 *IPTEK The Journal for Technology and Science* **28** 15-19
- [5] Solanki H U, Mankodi P C, Dwivedi R M and Nayak S R 2008 *Hydrobiologia* **612** 269-279
- [6] Moradi M and Moradi N 2020 *Marine Pollution Bulletin* **161** 111728
- [7] Solanki H U, Dwivedi R M, Nayak S R, Jadeja J V, Thakar D B, Dave H B and Patel M I 2001 *Indian J. Mar. Sci.* **30** 132-138.
- [8] Baldacci A, Corsini G, Grasso R, Manzella G, Allen J T, Cipollini P, Guymer T H and Snaith H M 2001 *J. Mar. Syst.* **29** 293-311
- [9] Nezlin N P and DiGiorgio P M 2005 *Cont. Shelf Res.* **25** 1692-1711
- [10] Longhurst A 1993 *Deep-Sea Res* **40** 2145–2165
- [11] Susanto R D, Moore II T S and Marra J 2006 *Geochem. Geophys. Geosy.* **7** 1-16
- [12] Ishizaka J 2010 *Proceedings of International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science XXXVIII Part 8, Kyoto, Japan*
- [13] Kuswanto T D, Syamsuddin ML and Sunarto 2017 *Journal of Fisheries and Marine Affairs* **VIII** 90-102
- [14] O'Reilly J E, Maritorena S, Mitchell B G, Siegel D A, Carder K L, Garver S A, Kahru M and McClain C 1998 *Journal Geophys Researcher* **103** 24937–24953
- [15] Prakash S and Ramesh R 2007 *Current Science* **92** 667-671
- [16] Greene W H 2018 *Econometric Analysis 8th edition* (New York: Pearson)
- [17] Zhang G P 2003 *Neurocomputing* **50** 159-175
- [18] Pankratz A 1983 *Forecasting with Univariate Box-Jenkins Model: Concepts and Cases* (New York: John Wiley & Sons)
- [19] Box G E P, Jenkins G M and Reinsel G C 1994 *Time Series Analysis: Forecasting and Control, 3rd edition* (Englewood Cliffs: Prentice-Hall).
- [20] Montgomery D C, Runger G C and Hubele N F 2011 *Engineering Statistics 5th edition* (Hoboken: John Wiley & Sons)
- [21] Holmes A, Illowsky B and Dean S 2017 *Introductory Business Statistics* (Houston: OpenStax)
- [22] Takeshige A, Takahashi T, Nakata H and Kimura S 2013 *Continental Shelf Research* **66** 73-82
- [23] Putra E, Gaol J L and Siregar V P 2012 *Journal of Fisheries and Marine Technology* **3** 1-10
- [24] Agung N A, Zainuri M, Wirasatriya A, Maslukah L, Subardjo P, Suryosaputro A A D and Handoyo G 2018 *Oceanographic Bulletin Marina* **7** 67-74
- [25] Chen H H, Tang R, Zhang H R, Yu Y, Wang Y 2020 *J. Vis. Exp.* **160** e61172
- [26] Kumar G S, Prakash S, Ravichandran R and Narayana A C 2016 *Remote Sensing Letters* **7** 1093–1101
- [27] Goes J I, Thoppil P G, Gomes H R and Fasullo J T 2005 *Science* **308** 545–547